Limit Theorems for Small Worlds *

Arkadiusz Szydłowski†

University of Kent

December 15, 2023

Abstract

I obtain central limit theorems for data coming from a sparse, small world network, i.e. a network with limited maximal degree and a relatively small (but growing) diameter, properties encountered in many social and economic networks. I do not impose restrictions on the strength of dependence between connected nodes and only assume that non-connected nodes are statistically independent. The limit theorems hold conditionally on the network evolution and are obtained by combining the bound on the possible number of connected nodes from algebraic graph theory and limit theorems for m-dependent time series. I consider both means of node- and edge-specific characteristics.

JEL: C10, C45

Keywords: Networks, CLT, m-dependence, Nonstationary data, Wild cluster bootstrap

1 Introduction

Many economic and social networks on top of being sparse exhibit a small-world property (Watts & Strogatz (1998)), namely that the network distance between each pair of connected nodes is

*I would like to thank Bristol Econometric Study Group participants, Vadim Marmer, Nizar Allouch and Alfred Duncan for useful comments.

†School of Economics, University of Kent, Canterbury, CT2 7PE, UK. *E-mail address*: a.szydlowski@kent.ac.uk

small compared to the number of nodes in the network.[1] Formally, a small-world network has a diameter proportional to $\log N$, where N is the number of nodes. We combine this restriction and boundedness of the maximal degree of a node (i.e. sparsity) and ask a question if one can obtain a central limit theorem (CLT) for a network data where observations from connected nodes (through any path) can be arbitrarily dependent and only unconnected nodes are statistically independent.

The answer to our main question is affirmative under an additional assumption restricting the size of the diameter relative to the maximal degree for any given N, or, put differently, restricting the constant of proportionality relating the diameter to $\log N$. We proceed with suggesting estimators for the variance of the sample mean and investigate their performance in Monte Carlo simulations. Although a simple block-variance estimator leads to considerable undercoverage of the resulting confidence interval, the performance of a wild cluster bootstrap procedure is very promising and we recommend using it for practical application of our results to hypothesis testing and building confidence intervals.

We proceed with our analysis conditional on network evolution, thus we do not include uncertainty coming from network formation. Hence, our results apply to stable networks with network-mediated dependence as the main source of dependence. An example is a long-term friendship network where we are interested in labour market outcomes. These outcomes are likely to have been affected by network interactions (e.g. referrals) and our results suggest how to proceed with inference on means of such outcomes.

In addition to node-specific means we also consider CLTs for edge-specific characteristics, where we distinguish between flows, i.e. purely characteristics of edges, and contrasts, i.e. functions of characteristics of nodes involved in an edge. We show that CLTs for edge-specific means hold under strengthening of conditions needed for node-specific means. In particular, we require stationarity to justify a CLT for means of contrasts.

Kojevnikov et al. (2021) provide a CLT for node-specific means assuming weak dependence in the form of ψ-dependence (see also Leung & Moon (2023)). They do not condition on the observed network in their analysis and provide some primitive conditions and examples of network formation processes consistent with ψ-dependence. They propose a HAC-type variance estimator. Leung

[1] The "small-world" property often also includes the characteristic that the network graph is much more clustered than a random graph (see e.g. Definition 4.1.3 in Watts (1999)). However, the latter property is not useful for the purpose of providing CLT in our context so we do not discuss it.

(2023) shows, however, that for many networks a cluster-robust inference may perform better. We find that a wild cluster bootstrap works quite well in our setup. Similarly to our paper Ogburn et al. (2022) provide a CLT conditional on the network formation process but only allow dependence up to friends-of-friends, whereas we allow for any connected nodes (via any path) to be dependent, thus generalising their results.

There is a large literature on obtaining limit theorems with spatial networks (see Jenish & Prucha (2012), Kuersteiner & Prucha (2013), Kuersteiner (2019) among others). Although many social and economic phenomena could be modelled using these networks, most social and economic networks have relatively high clustering coefficients, which means common presence of cliques (see Jackson (2008)). But Kojevnikov et al. (2021) demonstrate that spatial networks have limitations in terms of accommodating nontrivial presence of cliques, thus restricting their usefulness in modelling observed networks.

2 Main idea

Let $\{Y_1, \ldots, Y_N\}$ denote mean zero random variables corresponding to nodes in a network G_N. We are interested in the central limit theorem for the sample mean:

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

assuming that only nodes not connected through any network path have statistically independent Y_i's and dependence between the remaining observations is not restricted.

The proof of the main result combines the bound on the maximal number of nodes in a sparse network with a given maximal degree and diameter (Pineda-Villavicencio & Wood (2015)) with the central limit theorem for m-dependent random variables where $m \to \infty$ (Romano & Wolf (2000)). The key to the application of the latter result is showing that:

$$Var\left(\sum_{i=1}^{N} Y_i\right) = \sum_{i=1}^{N} Var(Y_i) + \sum_{i=1}^{N} \sum_{j \neq i} Cov(Y_i, Y_j)$$

is of order $O(N^{1+\alpha})$ where $0 \leq \alpha < 1$.

For simplicity assume that the network contains one giant component of maximal size N_C, denoted by g_N and the remaining components are of order $O(1)$, i.e. they do not grow with N. The case of multiple "large" components where the number of components grows with N follows similar arguments and is analysed in detail below. In this setup we have:

$$Var\left(\sum_{i=1}^{N} Y_i\right) = \sum_{i \in g_N} Var(Y_i) + \sum_{i \in g_N} \sum_{j \in g_N, j \neq i} Cov(Y_i, Y_j) + O(N),$$

the middle term is of order $N_C(N_C - 1)$ under our assumptions and this is the dominant term. Now the result from the algebraic graph theory bounds N_C by a quantity of order $d_{max}^{\Delta_N}$, where d_{max} denotes maximal degree and Δ_N denotes network's diameter. Therefore, it remains to provide conditions under which this upper bound is of order $O(N^{1+\alpha})$ and make sure that such α satisfies conditions of Romano & Wolf (2000).

Since our proof involves showing that N_C is of order smaller than N, the results do not apply to networks with a giant component involving almost all nodes or, alternatively, networks with giant component growing at rate N. This happens, for example, in the Facebook network where 99.91% of individuals belong to the largest connected component (Ugander et al. (2011)). Still, our limit theorems would apply if the giant component is of order arbitrarily smaller than N or if one can reasonably divide the giant component into statistically independent sub-components satisfying this condition (e.g. the links between some Facebook groups do not generate any cross-traffic, thus the groups can be viewed as independent). Finally, we note that if one is willing to assume weak dependence between connected nodes the results of Kojevnikov et al. (2021) would apply to such networks.

Recently, for a related problem, Kojevnikov & Song (2023) showed that consistent estimation of mean in clustered samples. without intra-cluster dependence restrictions, requires presence of at least two large clusters, which implies that the largest cluster has to be of order smaller than $O(N)$. Thus, our findings are in line with that result.

3 Central limit theorems for small-world networks

Recall that $d_{max} = \max_{i \in \mathcal{N}_N} d_i$ is the maximal degree in network G_N, where $\mathcal{N}_N = \{1, 2, \ldots, N\}$. Define $l(i,j)$ to be the network distance on the shortest path between i and j and set $l(i,j) = \infty$ if i and j are not connected by any network path. The diameter of network G_N is now formally defined as $\Delta_N = \max_{i,j \in \mathcal{N}_N : l(i,j) < \infty} l(i,j)$. All our results hold conditionally on network evolution $\{G_N\}_{N=1}^{\infty}$ and, thus, take network formation as given.

3.1 Node specific means

Consider the sample mean $\overline{Y}$ defined above and make the following assumptions:

Assumption 1. *(a)* $d_{max} \geq 2, d_{\max} = O(1)$.

(b) $\Delta_N \leq \log_a(bN)$ *for some constants* $a > 1, b > 0$.

(c) $l(i,j) = \infty$ *implies* $Y_i \perp Y_j$.

(d) $E|Y_i|^{2+\delta}$ *is bounded for all* i *and* $\delta > 0$.

(e) There exists $0 \leq \gamma < 1$ *such that:*

$$\log_{d_{max}-1}(a) > 1 + (1-\gamma)\left(1 + \frac{2}{\delta}\right).$$

Assumption 1(a) imposes sparsity of the network. The lower bound on the maximal degree has a technical nature and rules out only an (uninteresting) case of networks formed only of connected pairs of nodes. Note that sparsity is often imposed as a condition on average degree, namely $1/N \sum_{i=1}^{N} d_i = O(1)$, which is implied by our condition. However, for all practical purposes both formulations can be seen as equivalent. Assumption 1(b) imposes the small-world property, namely that the diameter of the network is (at most) proportional to $\log N$ and part (c) states that unconnected nodes are statistically independent. The crucial condition (e) restricts the scaling factor for the network diameter relative to the maximal degree and requires that for given N the diameter cannot be too large relative to the maximal degree (in other words, $1/\log(a)$ cannot be too large). We note that this condition is sufficient for obtaining a CLT, but need not be necessary. If all moments of Y exist, i.e. $\delta = \infty$, then $1 + (1-\gamma)\left(1 + \frac{2}{\delta}\right)$ can be arbitrarily close to one, which

implies that the size of the largest network component can grow at a rate arbitrarily smaller than N (see the proof of Theorem 1 in the Appendix).

Theorem 1. *Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of mean zero random variables and define $B_N^2 = Var(\sqrt{N}\overline{Y})$. Under Assumption 1:*

$$\frac{\sqrt{N}\overline{Y}}{B_N} \to^D N(0,1)$$

as $N \to \infty$ (conditionally on network evolution).

The proof of this CLT is given in the Appendix and follows the argument outlined in the previous section. Note that the presence of the factor $\sqrt{N}$ in the statement of theorem does not imply that we obtain a square root rate of convergence as in general B_N will not be $O(1)$. The result can be restated as a result conditional on common shocks affecting all the nodes in the network just as in Kojevnikov et al. (2021), and would then apply to networks where there is some dependence between unconnected nodes and the dependence can be modelled through observables.

4 Variance estimation

In this section we suggest estimators of the variance B_N^2 that can be used for inference together with Theorem 1. Since Y_i's have zero mean:

$$B_N^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E(Y_iY_j)\mathbb{1}\{l(i,j) < \infty\}$$

thus a natural estimator arises:

$$\hat{B}_N^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_iY_j\mathbb{1}\{l(i,j) < \infty\}.$$

Note that this estimator can be viewed as a block-variance estimator where the blocks correspond to different unconnected components of the network and grow in size with N.

In order to show consistency of this estimator we need a strengthening of Assumption 1(e):

Assumption 1(e)'. *There exists* $0 \leq \gamma < 1$ *such that:*

$$\log_{d_{max}-1}(a) > \max\left\{1 + (1-\gamma)\left(1 + \frac{2}{\delta}\right), 3\right\}.$$

Recall that Assumption 1(e) potentially allowed the size of the largest connected component to grow at rate arbitrarily close to N. Again, this is in line with findings in Kojevnikov & Song (2023) for clustered samples. This assumption implies that the maximal number of nodes in a connected component cannot grow faster than $N^{1/3}$ so puts a non-trivial restriction on the network architecture. They show that one requires much stricter conditions for variance estimation then for consistent discrimination of the mean.[2]

Theorem 2. *Let* $\{Y_i\}_{i=1}^{\infty}$ *be a sequence of mean zero random variables, Assumptions 1* (a)-(d) *and 1(e)' hold and additionally assume that* $E|Y_i|^4$ *is bounded for all* i*. Then:*

$$Var(\hat{B}_N - B_N) \to 0.$$

as $N \to \infty$ *(conditionally on network evolution).*

The theorem implies consistency of the proposed estimator. As the estimator only uses cross-products corresponding to observations in the same network component (alternatively, block) we coin it the block-variance estimator.

Although consistent, this estimator does not work well in practice if there is a lot of dependence between Y_i's. In our Monte Carlo simulations we show that a confidence interval using the block-variance estimator severely undercovers even for the sample size $N = 10000$ when there is strong dependence between observations belonging to the same network component. This is in line with simulations for a related HAC estimator in Kojevnikov et al. (2021) when the "autoregressive" parameter is close to 0.5.

As our setup is similar to the problem of estimating variance with a few large and growing in size clusters (Cameron et al. (2008)), as an alternative to the block-variance estimator we consider the wild clustered bootstrap and find that it performs much better in our Monte Carlo simulations.

[2]Note that the Hansen and Lee condition (Hansen & Lee (2019)) that they require for consistent estimation of variance is satisfied with clusters of size $N^{1/3}$.

Let $c = 1, \ldots, C_N$ enumerate separate components of network G_N. The bootstrap procedure is as follows:

1. For each connected component draw $v_c = -1$ or 1 with probability 1/2.
2. Resample Y_i's with replacement within components and obtain $\{Y_i^*\}_{i=1}^N$.
3. Calculate $\overline{Y}^* = \frac{1}{N}\sum_{i=1}^N Y_i^* v_{c(i)}$ where $c(i)$ denotes the component that i belongs to.
4. Estimate B_N^2 by variance of $\sqrt{N}\overline{Y}^*$ across bootstrap samples.

As an alternative one may consider randomisation tests of Canay et al. (2017).

5 Means of edge-specific characteristics

In this section we provide limit theorems for means of characteristics of edges between nodes. Applications include means of input-output flows in production networks (see e.g. Acemoglu et al. (2012)) or mean difference in socio-economic status between individuals belonging to the same local community (see e.g. Chetty et al. (2022)). Note that the edge characteristics in these two examples have a different structure – in the former they are nonparametric functions of a node pair (i, j) ("flows") whereas in the latter they are known functions of characteristics of a node (i, j) involved in an edge ("contrasts"). These differences lead to distinct analysis, in particular CLT for contrasts requires stronger conditions.

5.1 Flows

Let Y_{ij} denote the characteristic of an edge between nodes i and j and define the edge-specific mean as:

$$\overline{Y}_f = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} Y_{ij}.$$

where $Y_{ij} = 0$ if nodes i and j are not connected (by any path). Similarly to node-specific means we will assume that flows in separate network components are statistically independent:

Assumption 1(c)'. *$l(i, k) = \infty$ implies $Y_{ij} \perp Y_{kl}$.*

As there is no structure imposed on dependence within network components, we may have Y_{ij} and Y_{kl} arbitrarily correlated when (i, j, k, l) belong to a connected network component. Thus, one can analyse this case similarly to the node-specific case with a difference that now each component contains at most $N_C(N_C - 1)$ elements instead of N_C elements in the latter case.

Theorem 3. *Let* $\{Y_{ij}\}_{i,j=1}^{\infty}$ *be a sequence of mean zero random variables and define* $B_{N,f}^2 = Var(\sqrt{N(N-1)}\overline{Y}_f)$. *Under Assumptions 1* (a), (b), (d) *(with* Y_{ij} *replacing* Y_i*),* (e) *and 1(c)' :*

$$\frac{\sqrt{N(N-1)}\overline{Y}_f}{B_{N,f}} \to^D N(0,1)$$

as $N \to \infty$ *(conditionally on network evolution).*

Theorem 3 can be used for inference once an estimator of $B_{N,f}$ is available. One would expect that an analogous estimator to the block-variance estimator or a wild cluster bootstrap described in Section 4 would work, though we do not provide a formal result.

5.2 Contrasts

Let h be a symmetric function and define the edge-specific mean as:[3]

$$\overline{Y}_c = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} h(Y_i, Y_j) \mathbb{1}\{l(i,j) < \infty\}.$$

A leading example would be $h(Y_i, Y_j) = |Y_j - Y_i|$ with Y_i denoting a measure of socio-economic status like income (Chetty et al. (2022)), in which case the statistic would measure average differences in income among neighbourhoods ("economic connectedness") and our results would provide a starting point for conducting inference which takes into account network-dependence between connected units. We point out that, when there is non-negligible dependence between connected individuals, even large sample sizes may not guarantee statistical significance of the findings as in such case the "effective" sample size may be small.[4]

[3]The definition could be extended to functions of characteristics of triples, quadruples etc. of nodes, which can be used to study clique characteristics. The treatment of such statistics would follow similar lines. Hence, for the sake of exposition, we do not analyse them in detail.

[4]In Appendix F we analyze a version of the statistic $\overline{Y}_c = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} h(Y_i, Y_j)$, which allows contrasts between nodes belonging to different components to be nonzero. As this enriches the dependence structure between

Under stationarity of Y_i we get the following Hoeffding decomposition:

$$\frac{\sqrt{N}\overline{Y}_c}{B_{N,c}} = B_{N,c}^{-1}\frac{2}{\sqrt{N}}\sum_{i=1}^{N} h_1(Y_i)\frac{N_c(i)-1}{N-1} + B_{N,c}^{-1}\frac{\sqrt{N}}{N(N-1)}\sum_{i\neq j} h_2(Y_i,Y_j)\mathbb{1}\{l(i,j)<\infty\},$$

where $h_1(y) = E_Y[h(y,Y)], h_2(y_1,y_2) = h(y_1,y_2) - h_1(y_1) - h_1(y_2)$ and $N_c(i)$ denotes the number of nodes in the component to which i belongs. In order to obtain a CLT we need to show that the second term in the decomposition converges to zero in probability and show that a CLT holds for the triangular array $\left\{h_1(Y_i)\frac{N_c(i)-1}{N-1}\right\}$, where the latter result follows from the CLT in Romano & Wolf (2000).

Theorem 4. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary sequence and $E[h(Y_i,Y_j)] = 0$ for all i,j. Define $B_{N,c}^2 = Var\left(\frac{2}{\sqrt{N}}\sum_{i=1}^{N} h_1(Y_i)\frac{N_c(i)-1}{N-1}\right)$. Under Assumption 1 with part* (d) *replaced by: $E|h(Y_i,Y_j)|^{2+\delta}$ is bounded for all i,j and $\delta > 0$, we have:*

$$\frac{\sqrt{N}\overline{Y}_c}{B_{N,c}} \to^D N(0,1)$$

as $N \to \infty$ (conditionally on network evolution).

Compared to the previous results, this theorem imposes additional assumption of stationarity, which is needed for the Hoeffding decomposition, but otherwise uses very similar conditions.

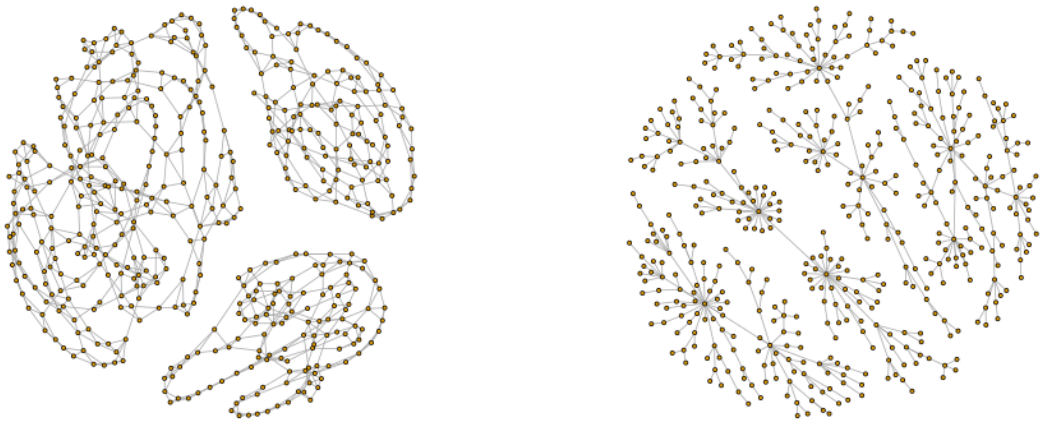
6 Monte Carlo simulations

We consider two network generating algorithms: the Watts-Strogatz small world (SW) model (Watts & Strogatz (1998)) and the Barabási-Albert (BA) preferential attachment model (Barabási & Albert (1999)). The first model generates networks with diameters proportional to $\log N$ whereas the second model produces diameters proportional to $\log N$ or $\log N/\log\log N$ depending on the parameters (Bollobás & Riordan (2004)). For most parameter values the BA model implies that the maximal degree of a node grows with N, thus we further "prune" the graph to make sure that the maximal degree is stable: (1) we start with a node with maximal degree and randomly erase superfluous edges, (2) check if maximal degree satisfies the imposed bound, (3) if not, we go back to step (1)

the h's we need much stronger conditions in order to provide a CLT there.

and repeat the procedure.

Figure 1: Examples of Monte Carlo designs, SW model (left) and BA model (right), $N = 500$.



Let $c(i)$ denote the network component containing node i and $N_c(i)$ denote the number of connected nodes in this component. The data is generated from a process similar to Kojevnikov et al. (2021):

$$Y_i = \frac{1}{\sqrt{N_c(i) - 1}} \sum_{j \neq i, j \in c(i)} \varepsilon_j$$

where ε_j's are i.i.d., drawn from a standardised uniform distribution.[5] In other words, node i's outcome is equal to the average of ε's of all the nodes that i is connected to, which implies strong dependence between outcomes belonging to the same network component. In terms of the architecture of the network, we start with four connected components for $N = 500$ and add one component for each increase in sample size above that, hence ending up with seven components for $N = 10000$. We perform 1000 MC repetitions and use 1000 replications for bootstrap procedures. Figure 1 shows two examples of networks generated by SW and BA models.

We consider coverage of confidence intervals built using known variance ("oracle") , block-variance estimator $\hat{B}_N$ ("estim.") and wild cluster bootstrap ("boot.") introduced in Section 4.

[5]If the components were equal-sized with N_c nodes each, $\overline{Y} = \sqrt{N_c - 1}\frac{1}{N}\sum_i \varepsilon_i$, so the CLT does not follow by construction as $N_c \to \infty$.

6.1 Node-specific means

Table 1 contains the simulation result for means of node-specific characteristics. The BA model is parametrised by: m – the number of edges added in each step of building the graph, z_a - appeal of nodes that do not have any connections. The algorithm of building the SW model starts with a circle (or, more generally, lattice) graph and "rewires" some of the connections between neighbouring nodes to some more distant nodes, thus is parametrised by: p - probability of rewiring an edge, k - the number of edges per vertex in the initial circle graph. Different values of these parameters produce graphs with different maximal degrees and diameters.

Table 1: Simulated coverage, node-specific means, 95% level

	BA model							SW model						
					Coverage							Coverage		
N	m	z_a	d_{max}	Δ_N	oracle	estim.	boot.	p	k	d_{max}	Δ_N	oracle	estim.	boot.
500	1	0	10	12	0.954	0.844	0.988	0.05	2	6	12	0.958	0.814	0.998
1000	1	0	10	11	0.964	0.930	0.954	0.05	2	6	15	0.950	0.854	0.975
5000	1	0	10	14	0.938	0.941	0.954	0.05	2	7	20	0.949	0.860	0.978
10000	1	0	10	16	0.952	0.950	0.942	0.05	2	7	20	0.960	0.897	0.972
500	1	1	10	16	0.954	0.869	0.983	0.05	5	13	5	0.949	0.836	0.994
1000	1	1	10	16	0.947	0.870	0.968	0.05	5	13	6	0.932	0.862	0.987
5000	1	1	10	18	0.957	0.928	0.955	0.05	5	14	7	0.954	0.876	0.965
10000	1	1	10	20	0.956	0.937	0.956	0.05	5	15	8	0.954	0.893	0.968
500	1	2	10	15	0.942	0.850	0.979	0.05	10	24	3	0.943	0.811	0.998
1000	1	2	10	17	0.955	0.869	0.951	0.05	10	24	4	0.946	0.828	0.983
5000	1	2	10	25	0.950	0.923	0.969	0.05	10	26	5	0.945	0.858	0.971
10000	1	2	10	21	0.941	0.920	0.968	0.05	10	26	5	0.941	0.904	0.957
500	2	0	20	6	0.959	0.818	0.994	0.10	2	8	9	0.947	0.815	0.995
1000	2	0	20	7	0.948	0.847	0.977	0.10	2	7	10	0.954	0.851	0.986
5000	2	0	20	10	0.956	0.876	0.967	0.10	2	8	14	0.951	0.862	0.968
10000	2	0	20	10	0.951	0.860	0.958	0.10	2	9	15	0.957	0.886	0.969
500	2	1	20	7	0.953	0.801	0.999	0.10	5	14	4	0.951	0.803	0.997
1000	2	1	20	7	0.950	0.831	0.980	0.10	5	14	5	0.943	0.836	0.970
5000	2	1	20	10	0.959	0.879	0.971	0.10	5	16	6	0.954	0.857	0.962
10000	2	1	20	10	0.948	0.875	0.970	0.10	5	16	7	0.961	0.883	0.963
500	2	2	20	6	0.951	0.809	0.995	0.10	10	25	3	0.953	0.800	0.995
1000	2	2	20	7	0.952	0.854	0.983	0.10	10	25	4	0.942	0.855	0.975
5000	2	2	20	9	0.955	0.884	0.968	0.10	10	28	4	0.946	0.874	0.975
10000	2	2	20	10	0.955	0.888	0.974	0.10	10	30	5	0.951	0.873	0.963

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. "Oracle" – known variance, "estim." – variance estimator $\hat{B}_N$, "boot." – wild cluster bootstrap.

When we use the known variance the coverage is close to the nominal 95% level across the designs

and parameter values, thus confirming that the CLT holds for small world networks. However, once we use the estimated variance $\hat{B}_N^2$ the coverage deteriorates substantially, with values somehow close to the nominal values only in the three top left panels of Table 1 for which the networks are pretty sparse with small degree and large diameter. This shows the difficulty of precisely estimating the variance with strong dependence between observations in a network setting, a phenomenon also occurring in Kojevnikov et al. (2021) (see their simulation results with large values of the "autoregressive" parameter γ).[6]

Interestingly, besides the small sample size $N = 500$, the wild cluster bootstrap works reasonably well with coverage values above the nominal 95% across all designs. However, we also see that it provides rather conservative inference. Still the overcoverage with the wild bootstrap is of much smaller magnitude than the undercoverage with the estimated variance, thus we recommend using the wild cluster bootstrap for data coming from sparse small world networks.

6.2 Edge-specific means

As the case of means of flows discussed in Section 5.1 is very similar to the case of node-specific means, we only run simulations for the means of contrasts. We use the same designs as above with a difference that we draw ε from a normal distribution and set:[7]

$$h(Y_i, Y_j) = |Y_j - Y_i|.$$

As variance estimation in the U-statistic setup is more involved than with simple means we do not provide variance estimators above, thus we only provide coverage values with known variance.

Table 2 shows that for all specifications of the network formation model the coverage probabilities are close to 95% even for $N = 500$, which is in line with our CLT in Theorem 5.

[6]Kojevnikov et al. (2021) use a HAC estimator with kernel weighting but the idea behind our $\hat{B}_N$ and their estimator is similar.

[7]We have also ran simulations with uniform errors, just as in the previous section, and the results are very similar. See Appendix G.

Table 2: Simulated coverage, edge-specific means (contrasts), 95% level

	BA model					SW model				
N	m	z_a	d_{max}	Δ_N	Coverage oracle	p	k	d_{max}	Δ_N	Coverage oracle
500	1	0	10	10	0.961	0.05	2	6	15	0.964
1000	1	0	10	11	0.958	0.05	2	7	15	0.964
5000	1	0	10	13	0.958	0.05	2	7	19	0.968
10000	1	0	10	14	0.939	0.05	2	8	21	0.964
500	1	1	10	16	0.961	0.05	5	14	5	0.947
1000	1	1	10	16	0.949	0.05	5	14	6	0.956
5000	1	1	10	23	0.967	0.05	5	14	7	0.958
10000	1	1	10	19	0.958	0.05	5	14	8	0.962
500	1	2	10	14	0.953	0.05	10	24	4	0.965
1000	1	2	10	15	0.953	0.05	10	26	4	0.956
5000	1	2	10	18	0.944	0.05	10	26	5	0.965
10000	1	2	10	21	0.959	0.05	10	27	5	0.961
500	2	0	20	7	0.949	0.1	2	7	10	0.960
1000	2	0	20	8	0.977	0.1	2	8	10	0.973
5000	2	0	20	11	0.957	0.1	2	8	13	0.940
10000	2	0	20	11	0.949	0.1	2	8	16	0.946
500	2	1	20	7	0.958	0.1	5	15	4	0.968
1000	2	1	20	7	0.96	0.1	5	15	5	0.948
5000	2	1	20	9	0.964	0.1	5	16	6	0.972
10000	2	1	20	11	0.943	0.1	5	17	7	0.967
500	2	2	20	6	0.963	0.1	10	28	3	0.963
1000	2	2	20	7	0.963	0.1	10	26	4	0.959
5000	2	2	20	9	0.958	0.1	10	27	5	0.975
10000	2	2	20	9	0.954	0.1	10	29	5	0.959

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. "Oracle" – known variance.

7 Conclusion

Many social and economic networks are sparse and are small-world. We show that data coming from such networks satisfies a central limit theorem under the additional assumption restricting the constant of proportionality of the diameter to $\log N$, even without imposing weak dependence between connected nodes.

Our result can be seen as a "possibility" theorem showing that a CLT applies quite generally to network-dependent data with the largest component of size N^α where $\alpha < 1$, so imposing weak dependence conditions within connected components seem not strictly necessary here. We also provide some evidence that the wild cluster bootstrap can be used successfully (though, anti-conservatively) to estimate the variance in this strongly dependent network setting, thus going beyond the theoretical appeal of our results and showing its potential practicality in applications.

We consider a simple setup of undirected unweighted networks but the results should extend naturally to directed networks and networks in which we can assign (bounded) weights to covariances of characteristics between two connected nodes. If these weights would vanish or decrease sufficiently fast between large connected components, then one could potentially be able to extend our results to networks with a giant component of size $O(N)$, i.e. larger than allowed in our current setup.

Appendix

A Proofs

A.1 Main lemmas

Lemma 1. *(Romano, Wolf (2000)) Let $\{X_{iN}\}$ be a triangular array of mean zero random variables and suppose $X_{1N}, \ldots, X_{d_N N}$ is a sequence of m_N-dependent random variables. Define:*

$$S^2_{N,k,a} = Var\left(\sum_{i=a}^{a+k-1} X_{iN}\right),$$
$$S^2_N = Var\left(\sum_{i=1}^{N} X_{iN}\right).$$

Assume the following conditions hold. For some $\delta > 0$ and some $-1 \leq \gamma < 1$:

(a) $E|X_{iN}|^{2+\delta} \leq \Psi_N$ *for all* i,

(b) $S^2_{N,k,a}/k^{1+\gamma} \leq K_N$ *for all* a *and all* $k \geq m_N$,

(c) $S^2_N/(d_N m_N^{\gamma}) \geq L_N$,

(d) $K_N/L_N = O(1)$,

(e) $\Psi_N/L_N^{(2+\delta)/2} = O(1)$,

(f) $m_N^{1+(1-\gamma)(1+2/\delta)}/d_N \to 0$.

Then, $S_N^{-1}\left(\sum_{i=1}^N X_{iN}\right) \to^D N(0,1)$.

Lemma 2. *(Pineda-Villavicencio, Wood (2015)) Every graph with minimum degree d_{min}, maximum degree d_{max} and diameter Δ_N has at most $2d_{min}(d_{max}-1)^{\Delta_N - 1} + 1$ vertices.*

Lemma 3. *(Yoshihara (1976)) Let $\{Y_i\}_{i=1}^{\infty}$ be an absolutely regular process with regularity coefficients denoted by β. If there are $\delta, M > 0$ so that for all j:*

$$\int\int |h(Y_i, Y_j)|^{2+\delta} dF(Y_i) dF(Y_j) \leq M, \qquad \int\int |h(Y_i, Y_j)|^{2+\delta} dF(Y_i, Y_j) \leq M,$$

then there is a constant K, such that for $m = \max\{i_2 - i_1, i_4 - i_3\}$ where $i_1 \leq i_2 \leq i_3 \leq i_4$, the following inequality holds:

$$|E[h_2(Y_{i_1}, Y_{i_2})h_2(Y_{i_3}, Y_{i_4})]| \leq K\beta^{\delta/(2+\delta)}(m),$$

where $h_2(y_1, y_2) = h(y_1, y_2) - h_1(y_1) - h_1(y_2)$.

B Proof of Theorem 1

Let $C \in \mathcal{C}$ denote a connected component, $C(i)$ denote the component that i belongs to and C_N be the number of connected components in the network. Under Assumption 1(d) we have:

$$\begin{aligned} S_N^2 \equiv Var\left(\sum_{i=1}^N Y_i\right) &= \sum_{i=1}^N Var(Y_i) + \sum_{i=1}^N \sum_{j \neq i} Cov(Y_i, Y_j)\mathbb{1}\{l(i,j) \leq \Delta_N\} \\ &= O(N) + \sum_{C \in \mathcal{C}} \sum_{i \in C(i)} \sum_{j \neq i, j \in C(i)} Cov(Y_i, Y_j). \end{aligned}$$

Note that the second term is dominant and, defining N_c as the number of nodes in a component C, can be bounded:

$$\sum_{C \in \mathcal{C}} \sum_{i \in C(i)} \sum_{j \neq i, j \in C(i)} Cov(Y_i, Y_j) \leq M \sum_{c=1}^{C_N} N_c(N_c - 1) \leq M(C_N + 1)N_C(N_C - 1) = O(NN_C),$$

where N_C is the maximal number of nodes in a connected component allowed under our Assumptions. The last inequality follows from noticing that the sum $\sum_{c=1}^{C_N} N_c(N_c - 1)$ is maximised (subject to $\sum_{c=1}^{C_N} N_c = N$ and $N_c \leq N_C$) when the nodes are equally distributed among $C_N = N/N_C$ components.[8]

Assumption 1(e) implies $N_C = o(N)$ (see below). For N large enough, the worst-case dependence structure that gives a fastest growing S_N^2 is when we have $C_N = \frac{N}{N_C}$ components of maximal size N_C. Further, if the data is N_c-dependent where $N_c < N_C$ it is also N_C-dependent. Thus, without loss of generality we can assume that the data is generated from the worst-case dependence structure

[8]Note that having one giant component of maximal size N_C and many small components of limited size, like in Section 2, gives $S_N^2 = O(N_C^2) + O(N)$, which is of smaller order than the bound above.

and is N_C-dependent, which means $S_N^2 \cong \nu N N_C$ for some constant $\nu > 0$.[9]

It remains to verify conditions of Lemma 1 (with $d_N = N$). Part (a) holds with $\Psi_N = O(1)$ by Assumption 1(d). For condition (b) w.l.o.g. we can take $a = 1$ and order the nodes by components, i.e. number connected nodes by consecutive numbers. Then we have for $k \geq N_C$ and N sufficiently large:

$$S_{N,k,1}^2 \leq Var\left(\sum_{i=1}^{N_C} Y_i\right) + Var\left(\sum_{i=N_C+1}^{k-N_C} Y_i\right) \leq 2MN_C^2 = 2MN_C^{1+\gamma}N_C^{1-\gamma} \leq k^{1+\gamma}2MN_C^{1-\gamma}$$

which shows that condition (b) is satisfied with $K_N = 2MN_C^{1-\gamma}$. Furthermore, we have

$$S_N^2 \geq 0.5\nu N N_C = N N_C^{\gamma} 0.5\nu N_C^{1-\gamma}$$

so condition (c) of Lemma 1 is satisfied with $L_N = 0.5\nu N_C^{1-\gamma}$. Next Ψ_N, K_N and L_N satisfy conditions (d) and (e). For condition (f) we need:

$$\frac{N_C^{1+(1-\gamma)(1+2/\delta)}}{N} \to 0$$

which, again using the bound in Lemma 2 and taking $\log_{d_{max}-1}$ of the ratio above, is implied by

$$\log_a N\left(1 + (1-\gamma)\left(1 + \frac{2}{\delta}\right)\right) - \log_{d_{max}-1} N < 0$$

which can be rewritten as:

$$\log N \frac{\log(d_{max}-1)\left(1 + (1-\gamma)\left(1+\frac{2}{\delta}\right)\right) - \log a}{\log a \log(d_{max}-1)} < 0.$$

This inequality is implied by our Assumption 1(e).

Finally, noting that $S_N = \sqrt{N}B_N$ the result in Theorem 1 follows from Lemma 1.

[9] We can always group nodes into (a growing number of) components of size N_C where observations from different components are statistically independent.

C Proof of Theorem 2

We have:

$$
\begin{aligned}
N^2 Var(\hat{B}_N - B_N) &= E\left(\sum_{i=1}^{N}\sum_{j}(Y_iY_j - E(Y_iY_j))\mathbb{1}\{l(i,j)<\infty\}\right)^2 \\
&= \sum_{i=1}^{N}\sum_{j}\sum_{k}\sum_{l} E[(Y_iY_j - E(Y_iY_j))(Y_kY_l - E(Y_kY_l))]\mathbb{1}\{l(i,j)<\infty\}\mathbb{1}\{l(k,l)<\infty\} \\
&= \sum_{i=1}^{N}\sum_{j}\sum_{k}\sum_{l} Cov(Y_iY_j, Y_kY_l)\mathbb{1}\{l(i,j)<\infty\}\mathbb{1}\{l(k,l)<\infty\}. \qquad (1)
\end{aligned}
$$

But $\mathbb{1}\{l(i,j)<\infty\}\mathbb{1}\{l(k,l)<\infty\} = 0$ unless i and j belong to the same network component and same happens for k and l. But for such pairs of (i,j) and (k,l) we have $Cov(Y_iY_j, Y_kY_l) \neq 0$ only when (i,j) and (k,l) belong to the same network component (see Assumption 1(c)). Thus, we can rewrite (1) as:

$$
(1) = \sum_{C\in\mathcal{C}}\sum_{i\in C(i)}\sum_{j\in C(i)}\sum_{k\in C(i)}\sum_{l\in C(i)} Cov(Y_iY_j, Y_kY_l) \leq MNN_C^3
$$

because $Cov(Y_iY_j, Y_kY_l) \leq M$ for bounded M by the assumption that $E|Y_i|^4$ is bounded (in the statement of the theorem) and Cauchy-Schwartz inequality. The last inequality follows by the same reasoning as the one in the proof of Theorem 1.

Now Assumption 1(e)' implies $N_C = o(N^{1/3})$, which gives $Var(\hat{B}_N - B_N) \to 0$.

D Proof of Theorem 3

Note that we have N^2 observations in the sample and the data is N_C^2-dependent here. By the same reasoning as above we can bound:

$$
S_{N,f}^2 \equiv Var\left(\sum_{i=1}^{N}\sum_{j} Y_{ij}\right) \leq MN^2N_C(N_C-1)
$$

observing that the maximal number of correlated $Y_{ij}'s$ in a connected network component is $N_C(N_C-1)$ and, thus, we can have at most $C_N = N^2/N_C(N_C-1)$ "clusters" of correlated observations.

Now we can proceed with verifying conditions of Lemma 1 as in the proof of Theorem 1. In

particular, for condition (f) we need:

$$\frac{N_C^{2(1+(1-\gamma)(1+2/\delta))}}{N^2} \to 0$$

which is implied by our Assumption 1(e) (see above).

E Proof of Theorem 5

For a sequence of random variables $\{W_1, \ldots, W_N\}$ define the U_N operator as:

$$U_N h = \frac{1}{N(N-1)} \sum_{i \neq j} h(W_i, W_j) \mathbb{1}\{l(i,j) < \infty\}.$$

By Hoeffding decomposition:

$$\frac{\sqrt{N}\overline{Y}_c}{B_{N,c}} = B_{N,c}^{-1} \frac{2}{\sqrt{N}} \sum_{i=1}^{N} h_1(Y_i) \frac{N_c(i)-1}{N-1} + B_{N,c}^{-1} \sqrt{N} U_N h_2 \tag{2}$$

where $h_2(y_1, y_2) = h(y_1, y_2) - h_1(y_1) - h_1(y_2)$. Let us first show that $B_{N,c}^{-1}\sqrt{N}U_N h_2 = o_p(1)$.

Denoting $w_{iN} = (N_c(i)-1)/(N-1)$, by Cauchy-Schwartz inequality we can bound:

$$B_{N,c}^2 = \frac{4}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E[h_1(Y_i) h_1(Y_j)] w_{iN} w_{jN} \leq \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} E[h_1^2(Y_i) h_1^2(Y_j)]} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{iN}^2 w_{jN}^2} \leq M N_C = O(N_C)$$

(note that w_{iN}'s are nonstochastic conditional on the network). Again, as in the proof of Theorem 1 w.l.o.g. assume the worst-case dependence structure, i.e. $B_{N,c}^2 \cong \nu N_C$.

We have:

$$Var(\sqrt{N}U_N h_2) = \frac{1}{N(N-1)^2} \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k=1}^{N} \sum_{l \neq k} E[h_2(Y_i, Y_j) h_2(Y_k, Y_l)] \mathbb{1}\{l(i,j) < \infty\} \mathbb{1}\{l(k,l) < \infty\}.$$

The term under the sum is only nonzero if (i,j,k,l) belong to the same component (note that

$Eh_2(Y_i, Y_j) = 0$). Thus:

$$Var(\sqrt{N}U_N h_2) = \frac{1}{N(N-1)^2} \sum_{C\in\mathcal{C}} \sum_{i\in C} \sum_{j\in C, j\neq i} \sum_{k\in C} \sum_{l\in C, l\neq k} E[h_2(Y_i, Y_j)h_2(Y_k, Y_l)]$$
$$\leq \frac{1}{N(N-1)^2} MC_N N_C^2 (N_C - 1)^2 \simeq MN_C \left(\frac{N_C - 1}{N-1}\right)^2 = o(N_C),$$

which implies $\sqrt{N}U_N h_2 = o_p(N_C)$. Together with $B_{N,c}^2 \cong \nu N_C$ this implies that the second term in the decomposition in (2) converges to zero in probability.

Finally, the asymptotic normality of the first element in this decomposition follows from Lemma 1. Let us verify the conditions. Firstly, condition (a) follows with $\Psi_N = O(1)$ since $E|h(Y_i, Y_j)|^{2+\delta} < \infty$ implies $E|h_1(Y_i)|^{2+\delta} < \infty$ and $w_{iN} < 1$. Next, by Cauchy-Schwartz inequality $S_{N,k,1}^2 = Var(\sum_{i=1}^k h_1(Y_i)w_{iN}) = O(N_C^2)$ and $S_N^2 = Var(\sum_{i=1}^N h_1(Y_i)w_{iN}) = O(NN_C)$ so the remaining conditions for Lemma 1 follow by exactly the same argument as in the proof of Theorem 1.

F Additional result for contrasts

Let h be a symmetric function and define the edge-specific mean as:

$$\overline{Y}_c = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j\neq i} h(Y_i, Y_j)$$

In order to provide CLT for means of contrasts of this form we need to impose noticeably stronger conditions:

Assumption 2. *(a)* $\int\int |h(Y_i, Y_j)|^{2+\delta} dF(Y_i) dF(Y_j)$ *is finite for* $\delta > 0$.

(b) $\{Y_i\}_{i=1}^\infty$ *is a stationary, absolutely regular process with absolute regularity coefficients satisfying:* $\beta(N) = O\left(N^{-\epsilon\left(\frac{2+\delta}{\delta}\right)}\right)$ *for some* $\epsilon > 0$.

See e.g. Dehling & Wendler (2010) for the definition of absolute regularity. Absolute regularity is weaker than uniform mixing and stronger than strong mixing.

We need stationarity in order to use the Hoeffding decomposition:

$$\frac{\sqrt{N\overline{Y}_c}}{B_{N,c}} = B_{N,c}^{-1} \frac{2}{\sqrt{N}} \sum_{i=1}^N h_1(Y_i) + B_{N,c}^{-1} \sqrt{N} U_N h_2$$

where the operator U_N is defined below.

Next, in order to obtain a CLT we need to show that the second term in the decomposition converges to zero in probability. But the variance of $\sqrt{N}U_N h_2$ is of order $O(N_C)$ as all the covariances $Cov(h_2(Y_i, Y_j), h_2(Y_k, Y_l)), i \neq j, k \neq l$, for which at least one of (i, j) and at least one of (k, l) belongs to the same component are not zero. At the same time $B_{N,c}$ is of order $O(N_C)$ as well, thus preventing the last term in the decomposition to vanish with $N \to \infty$. This necessitates imposing stronger dependence condition within the components such that $Cov(h_2(Y_i, Y_j), h_2(Y_k, Y_l))$ vanishes as nodes i and k are further apart from each other in terms of the network distance (but still $l(i, k) < \infty$).

Theorem 5. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary sequence and $E[h(Y_i, Y_j)] = 0$ for all i, j. Define $B_{N,c}^2 = Var(2/\sqrt{N}\sum_{i=1}^{N} h_1(Y_i))$. Under Assumption 1 with part* (d) *replaced by: $E|h(Y_i, Y_j)|^{2+\delta}$ is bounded for all i, j and $\delta > 0$, and Assumption 2 we have:*

$$\frac{\sqrt{N}\overline{Y}_c}{B_{N,c}} \to^D N(0, 1)$$

as $N \to \infty$ (conditionally on network evolution).

The proof of this theorem uses the results on U-statistics with weakly dependent data (see e.g. Yoshihara (1976), Dehling (2006), Dehling & Wendler (2010)) in the context of m-dependent sequences when m is allowed to increase with sample size:

Proof. For a sequence of random variables $\{W_1, \ldots, W_N\}$ define the U_N operator as:

$$U_N h = \frac{1}{N(N-1)} \sum_{i \neq j} h(W_i, W_j).$$

By Hoeffding decomposition:

$$\frac{\sqrt{N}\overline{Y}_c}{B_{N,c}} = B_{N,c}^{-1} \frac{2}{\sqrt{N}} \sum_{i=1}^{N} h_1(Y_i) + B_{N,c}^{-1}\sqrt{N}U_N h_2$$

where $h_2(y_1, y_2) = h(y_1, y_2) - h_1(y_1) - h_1(y_2)$. The asymptotic normality of the first element in this decomposition follows from Theorem 1 noting that $E|h(Y_i, Y_j)|^{2+\delta} < \infty$ implies $E|h_1(Y_i)|^{2+\delta} < \infty$. So it remains to show that $B_{N,c}^{-1}\sqrt{N}U_N h_2 = o_p(1)$.

Order observations such that observations with neighbouring indices belong to the same network component and note that the maximal network distance between any two nodes is N. Since we have at most $C_N = N/N_C$ components of size N_C, using Lemma 3 with m replaced by N we can bound:

$$\begin{aligned} Var(\sqrt{N}U_N h_2) =& \frac{1}{N(N-1)^2} \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k=1}^{N} \sum_{l \neq k} E[h_2(Y_i, Y_j) h_2(Y_k, Y_l)] \\ & \leq \frac{C_N N_C^2 (N-1)^2 \beta^{\delta/(2+\delta)}(N)}{N(N-1)^2} = N_C \beta^{\delta/(2+\delta)}(N). \end{aligned}$$

But $\beta^{\delta/(2+\delta)}(N) = o(1)$ by our Assumption 2. Now, the argument leading to asymptotic normality of the first term in the Hoeffding decomposition (see proof of Theorem 1) implies that $B_{N,c} \geq \nu N_C$. Putting both results together we obtain $B_{N,c}^{-1}\sqrt{N}U_N h_2 = o_p(1)$. □

G Additional MC simulations

Table 3: Simulated coverage, edge-specific means (contrasts), uniform errors, 95% level

	BA model					SW model				
N	m	z_a	d_{max}	Δ_N	Coverage oracle	p	k	d_{max}	Δ_N	Coverage oracle
500	1	0	10	10	0.961	0.05	2	6	15	0.964
1000	1	0	10	11	0.958	0.05	2	7	15	0.964
5000	1	0	10	13	0.958	0.05	2	7	19	0.968
10000	1	0	10	14	0.939	0.05	2	8	21	0.964
500	1	1	10	16	0.961	0.05	5	14	5	0.947
1000	1	1	10	16	0.949	0.05	5	14	6	0.956
5000	1	1	10	23	0.967	0.05	5	14	7	0.958
10000	1	1	10	19	0.958	0.05	5	14	8	0.962
500	1	2	10	14	0.953	0.05	10	24	4	0.965
1000	1	2	10	15	0.953	0.05	10	26	4	0.956
5000	1	2	10	18	0.944	0.05	10	26	5	0.965
10000	1	2	10	21	0.959	0.05	10	27	5	0.961
500	2	0	20	7	0.949	0.1	2	7	10	0.960
1000	2	0	20	8	0.977	0.1	2	8	10	0.973
5000	2	0	20	11	0.957	0.1	2	8	13	0.940
10000	2	0	20	11	0.949	0.1	2	8	16	0.946
500	2	1	20	7	0.958	0.1	5	15	4	0.968
1000	2	1	20	7	0.96	0.1	5	15	5	0.948
5000	2	1	20	9	0.964	0.1	5	16	6	0.972
10000	2	1	20	11	0.943	0.1	5	17	7	0.967
500	2	2	20	6	0.963	0.1	10	28	3	0.963
1000	2	2	20	7	0.963	0.1	10	26	4	0.959
5000	2	2	20	9	0.958	0.1	10	27	5	0.975
10000	2	2	20	9	0.954	0.1	10	29	5	0.959

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. "Oracle" – known variance.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. (2012), 'The network origins of aggregate fluctuations', *Econometrica* **80**(5), 1977–2016.

Barabási, A.-L. & Albert, R. (1999), 'Emergence of scaling in random networks', *Science (American Association for the Advancement of Science)* **286**(5439), 509–512.

Bollobás, B. & Riordan, O. (2004), 'The diameter of a scale-free random graph', *Combinatorica* **24**(1), 5–34.

Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2008), 'Bootstrap-based improvements for inference with clustered errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Canay, I. A., Romano, J. P. & Shaikh, A. M. (2017), 'Randomization tests under an approximate symmetry assumption', *Econometrica* **85**(3), 1013–1030.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M. & Wernerfelt, N. (2022), 'Social capital i: measurement and associations with economic mobility', *Nature* **608**(7921), 108–121.

Dehling, H. (2006), Limit theorems for dependent U-statistics, *in* P. Bertail, P. Doukhan & P. Soulier, eds, 'Dependence in Probability and Statistics', Vol. 187, Springer New York, United States.

Dehling, H. & Wendler, M. (2010), 'Central limit theorem and the bootstrap for U-statistics of strongly mixing data', *Journal of Multivariate Analysis* **101**(1), 126–137.

Hansen, B. E. & Lee, S. (2019), 'Asymptotic theory for clustered samples', *Journal of Econometrics* **210**(2), 268–290.

Jackson, M. O. (2008), *Social and economic networks*, Princeton University Press, Princeton, N.J. ; Woodstock.

Jenish, N. & Prucha, I. R. (2012), 'On spatial processes and asymptotic inference under near-epoch dependence', *Journal of Econometrics* **170**(1), 178–190.

Kojevnikov, D., Marmer, V. & Song, K. (2021), 'Limit theorems for network dependent random variables', *Journal of Econometrics* **222**(2), 882–908.

Kojevnikov, D. & Song, K. (2023), 'Some impossibility results for inference with cluster dependence with large clusters', *Journal of Econometrics* **237**(2, Part A), 105524.

Kuersteiner, G. M. (2019), Limit theorems for data with network structure. Working paper.

Kuersteiner, G. M. & Prucha, I. R. (2013), 'Limit theory for panel data models with cross sectional dependence and sequential exogeneity', *Journal of Econometrics* **174**(2), 107–126.

Leung, M. P. (2023), 'Network cluster-robust inference', *Econometrica* **91**(2), 641–667.

Leung, M. P. & Moon, H. R. (2023), Normal approximation in large network models. Working Paper.

Ogburn, E. L., Sofrygin, O., Díaz, I. & van der Laan, M. J. (2022), 'Causal inference for social network data', *Journal of the American Statistical Association* **0**(0), 1–15.

Pineda-Villavicencio, G. & Wood, D. R. (2015), 'The degree-diameter problem for sparse graph classes', *The Electronic journal of combinatorics* **22**(2), 1–20.

Romano, J. P. & Wolf, M. (2000), 'A more general central limit theorem for m-dependent random variables with unbounded m', *Statistics & Probability Letters* **47**(2), 115–124.

Ugander, J., Karrer, B., Backstrom, L. & Marlow, C. (2011), The anatomy of the Facebook social graph. Working Paper.

Watts, D. J. (1999), *Small worlds : the dynamics of networks between order and randomness*, Princeton studies in complexity, Princeton University Press, Princeton, N.J.

Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of "small-world" networks', *Nature* **393**(6684), 440–442.

Yoshihara, K. (1976), 'Limiting behavior of U-statistics for stationary, absolutely regular processes', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **35**(3), 237–252.